Extraction of Tumours from MR Images of the Brain by Texture and Clustering

João Batista* and Richard Kitney

Imperial College, Electrical Engineering Dept.
Exhibition Road, SW7 2BT, London - UK

Abstract. The characterisation of tumours from Magnetic Resonance (MR) images of the brain is still a challenging task. In this paper we present an approach based on a K–Means clustering algorithm combined with textural feature information as opposed to intrinsic MR paramenters T1,T2 and PD. This is due to the fact that MR parameters may exhibit significant alterations in the presence of pathological conditions and, therefore lead to incorrect classification. We also address two important aspects of clustering: the selection of the optimum number of classes (Cluster Validity) and the most effective features (reduction of the feature space).

Index Terms: Tissue Characterisation, Texture Analysis, Clustering, Feature Reduction, Cluster Validity.

1 Introduction

Considerable work has been done on the characterisation of normal and abnormal human brain tissue based on quantitative Magnetic Resonance Imaging (MRI), ie the analysis and interpretation of its intrinsic parameters Proton Density (PD) and the relaxation times T1 and T2. It has also been shown that these parameters vary from tissue to tissue [Fos 84; Tay 86] but may exhibit significant alterations in the presence of pathological conditions, not allowing classification to be carried out exclusively on PD, T1 and T2 values.

The approach presented in this paper regards the classification of MRI brain tissues as a pattern recognition task [Bez 93], which combines features extracted from the original image – textures – with a classification procedure – clustering – to devise the classes, or tissues.

Texture analysis plays an important role in the segmentation of images which do not exhibit a clear boundary between different objects within the image. This can be observed in MR images of the brain, especially if oedema or tumour is present.

The main reason for using clustering, as opposed to supervised classifiers, is that the former does not require a training data set. The imaged physical values

in MRI vary from machine to machine due to different magnetic field values. Consequently, a training data set which is taken from a particular machine and then used to tune the classifier, may give misleading results if the classifier is applied to a different set of images – with the same physical characteristics as the training data set images – acquired from a different scanner. Moreover, an unsupervised classifier may reveal unnoticed structures within the image; also, no *a priori* number of classes needs to be defined.

We also discuss two important aspects of clustering: cluster validity and feature reduction. The former is concerned with the optimum number of clusters and the latter with redundant and/or undesirable features in the feature space.

2 Texture

When the discrimination of classes based on the grey level differences is ineffective, textural features can be considered. Texture is a neighbourhood property of an image and, therefore, it conveys properties such as fineness, coarseness, smoothness, granulation and randomness. Each of these nouns translates into some properties of the tonal primitives and the spatial interaction between pixels. As defined in [Goo 85], this interaction can be either deterministic or statistical.

For MR images, the nature of texture is typically statistical. The two methods described in this section aim at the characterisation of stochastic properties of the spatial distribution of the grey level in the image.

2.1 Spatial Grey Level Dependence Matrix (SGLDM)

The SGLDM describes the spatial distribution and spatial dependence among the grey tones in a local area based on the estimation of second order conditional probability density functions $f(i, j, d, \theta)$. Each of these functions are the probability of going from grey level i to grey level j separated by a distance d and aligned to the angle θ. The estimated values (which are grey–tone spatial–dependence frequencies) can be written in matrix form, the so–called co–occurrence matrices.

A number of texture features can be computed from the horizontal (0°), vertical (90°), left diagonal (45°) and right diagonal (135°) matrices. Haralick [Har 73] proposed 28 features extracted from 14 equations, but usually only 5 of them are used. They are angular second moment(ASM), contrast, correlation, inverse difference moment(IDM) and entropy.

2.2 Grey Level Run Length Method (GLRL)

The GLRL method is based on the calculation of the number of grey level runs with different lengths. A grey level run is a set of consecutive, co–linear pixels having the same grey level values. The length of the run is the number of pixels in the run.

Given an image and a direction (0°, 45°, 90°, 135°), a GLRL matrix can be computed. The GLRL matrix element (i, j) specifies the number of times the picture contains a run of length j, in the given direction, consisting of points having grey level values i, or within the grey level range i. Textural information can now be extracted from the GLRL matrices. A set of six functions used in

this work can be seen in [Loh 88]. They are second moment with respect to run length, first and second moments with respect to grey level, grey level and run length non-uniformity and sum of variance.

3 Clustering

The objective of cluster analysis is to separate the features into groups so that the members of any one group differ from each other as little as possible, according to a chosen criterion. The clustering used in the work which underlies this paper is based on the K-Means clustering algorithm of Coleman and Andrews [Col 79] and comprises three major steps: (a) initialisation, (b) distribution of feature vectors among existing clusters, (c) cluster validity.

(a) **Initialisation**: consists of splitting the input data set into two initial clusters. The input is given by a set of m feature vectors $v = (f_1, f_2, \ldots, f_n)$ where m is the number of pixels in the image, n is the feature vector dimension and $f_i (i = 1, \ldots, n)$ is a texture feature. The two initial clusters are created by computing the mean and variance over the m feature vectors. Cluster centres are then calculated to be evenly spaced on the diagonal of positive correlation of ± 1 standard deviation in the hyperspace of the feature set.

(b) **Distributing input among clusters**: Having established the new cluster centres, the M feature vectors must be assigned to the existing clusters until the algorithm converges. This is done by assigning each feature vector to the closest cluster centre through the Euclidean Distance [Fuk 72].

This step is repeated until all the cluster centres remain unchanged, ie until the algorithm has converged. At this point a new cluster can be devised by selecting amongst all clusters the feature vector which presents the largest distance from its cluster centre.

(c) **Cluster Validity**: One of the central points in clustering is to define the "correct" number of clusters for a given feature space. One possible approach is to obtain a measure of clustering quality represented by a parameter β derived from the within-cluster and the between-cluster scatter matrices [Fuk 72].

The within-cluster scatter matrix shows the scatter of samples about the cluster means and is expressed by:

$$S_w = \frac{1}{K} \sum_{k=1}^{K} \frac{1}{M_k} \sum_{x_i \in S_k} (x_i - \mu_k)(x_i - \mu_k)^T \tag{1}$$

where μ_k is the mean of the kth cluster; M_k is the number of elements in the kth cluster; x_i is an element in the kth cluster (the set of such elements given by S_k) and K is the total number of clusters.

The between-cluster scatter matrix shows the scatter of clusters in relation to the overall input set:

$$S_b = \frac{1}{K} \sum_{k=1}^{K} (\mu_k - \mu_o)(\mu_k - \mu_o)^T \tag{2}$$

where μ_o is the overall mean vector of the input set and μ_k is the mean vector of the kth cluster.

The β value which has produced the best results in this work is given by: $\beta_5 = \text{tr}S_b \cdot \text{tr}S_w$, where $\text{tr}(\bullet)$ indicates trace (sum of the diagonal elements of a matrix). Note that when the number of clusters is equal to 1, $\text{tr}S_w = \sigma^2$ (the variance of the input data set), $\text{tr}S_b = 0$ and $\beta_5 = 0$. When the number of clusters equal M (M is the total number of feature vectors), $\text{tr}S_w = 0$ and $\text{tr}S_b = \sigma^2$, hence, $\beta_5 = 0$. This measure is 0 at the limiting points of the clustering and greater than 0 in the interval. Therefore, it must convey at least one maximum value somewhere in the interval.

4 Reduction of Feature Space

Another fundamental problem in clustering is the selection of features which are most effective in producing an optimum class or cluster separability. This also reduces the computation time while improving the quality of the classification by discarding noisy, redundant and less useful features. The Bhattacharyya Distance (B–distance) is a measure of how much each feature contributes to the cluster separability for a pair of clusters (S_1, S_2) and is given by

$$B_n(S_i, S_j) = \frac{1}{4}\ln\left\{\frac{1}{4}\left(\frac{\sigma_i^2(n)}{\sigma_j^2(n)} + \frac{\sigma_j^2(n)}{\sigma_i^2(n)} + 2\right)\right\} + \frac{1}{4}\left\{\frac{(\mu_i(n) - \mu_j(n))^2}{\sigma_i^2(n) + \sigma_j^2(n)}\right\} \quad (3)$$

where n refers to the nth dimension of the feature space and σ_i and σ_j are variances of the ith and jth cluster data in dimension n, and μ_i and μ_j are their respective means.

The greater the differences in variance and mean for each pair, the larger the distances will be. The feature rejection criterion would be to retain those features which produce large B–distance measures. However, so that feature reduction can retain the best features, the input data set should be decorrelated (Hottelin Transform [Col 79]). That will rotate the features and orient them in the direction of those with higher eigenvector and, therefore, produce higher Bhattacharyya Distance. These features should then be retained and clustering performed again on the reduced feature set.

5 Application

We will now show the method applied to real MRI data. The data for this purpose is a pathological coronal T1-weighted image of the brain and is the 78th slice of a set of 124 images acquired from a 1.5 TESLA scanner, with TR=3500 and TE=5000. Seven textural measures have been selected using a 7×7 overlapping window: four from the GLRL method (first and second soment with respect to grey level, grey level non–uniformity and sum of variance) and three measures are functions from the SGLDM (contrast, correlation and inverse difference moment).

Figure 1 shows the normalised Bhattacharyya distance for each of the seven features. Both graphs show reasonably consistent behaviour of the distance measures as the number of clusters varies. Thus, selection based on this measure is a consistent procedure. The best of the rotated features (first moment with respect to grey level) was higher in Bhattacharyya distance than any other feature.
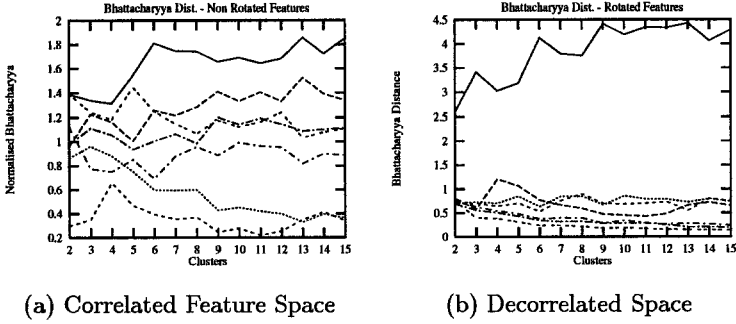
Bhattacharyya Dist. - Non Rotated Features

Bhattacharyya Dist. - Rotated Features

(a) Correlated Feature Space

(b) Decorrelated Space

Fig. 1. Bhattacharyya Distances - Feature Reduction

In order to identify the optimum number of clusters, the procedure is carried out with a fixed and excessive number of clusters: 15. Figure 2 plots the resulting β values against the number of clusters. Plot (a), shows 7 clusters as being the optimal value for all features. In plot (b), only the best feature has been selected and the optimum number of clusters was 5.
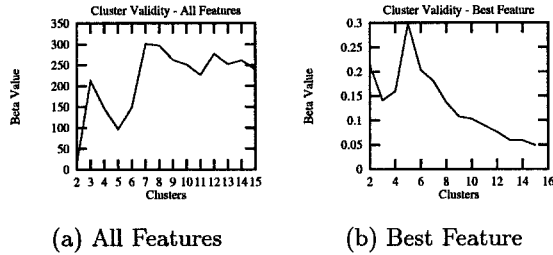
Cluster Validity - All Features

Cluster Validity - Best Feature

(a) All Features

(b) Best Feature

Fig. 2. Evaluation of the best number of clusters

At this point, the algorithm is again executed. It can be observed from figure 3 that the segmentation in (c) - although with misclassified regions - is more detailed than in (b). This shows that a reduction in features space has improved the classification. Also, it shows that a reduction in the number of clusters - from 7 in (b) to 5 in (c) - does not worsen the resulting classification. Note that in (d), the tumour has been partitioned in two cluster (a darker area, sorrounded by a lighter region). Finally, image (e) presents the classification for all seven features and 15 clusters. In (f), the isolated cluster corresponding to the extracted tumour is shown. Notice that the tumour is entirely extracted (with some misclassified regions though), despite the large pre–defined number of clusters.

6   Discussion and Conclusions

In this paper we have presented an approach to the extraction of tumours from MR images of the brain based on clustering and texture. The advantages of clustering and texture as features which convey important pathological information
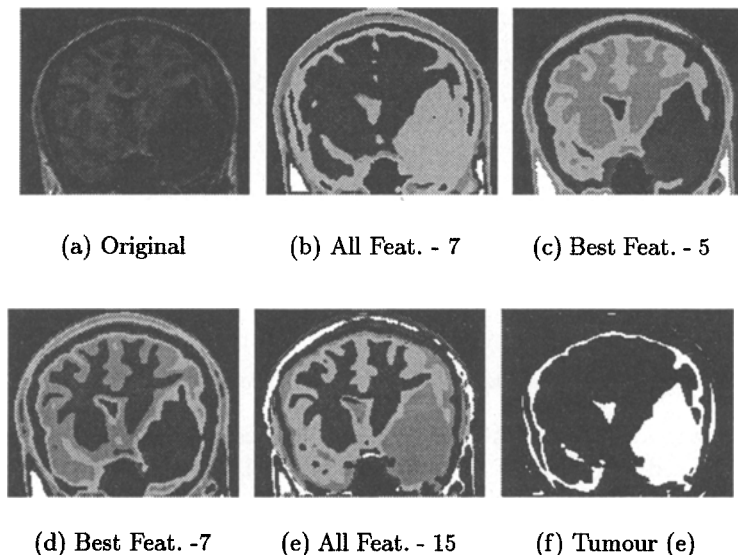
(a) Original   (b) All Feat. - 7   (c) Best Feat. - 5

(d) Best Feat. -7   (e) All Feat. - 15   (f) Tumour (e)

Fig. 3. Examples of tumour extraction by clustering

– as opposed to overlapping MRI parameters – have been described. The benefits of techniques for reducing the dimension of the feature space and estimating the appropiate number of clusters have also been discussed.

Finally, the segmentation procedure can be further improved by careful attention to window size and shape for feature acquisition, and the capacity to recognise the shape of regions in addition to their own texture.

References

[Bez 93] Bezdek, J.C.; Hall, L.O.; Clarke, L.P. - *Review of mr image segmentation tecniques using pattern recognition*. Medical Physics. 20(4):1033-1048, 1993.

[Col 79] Coleman,G.B.; Andrews,H.C. - *Image Segmentation by Clustering*. Proc. IEEE. 67(5):773-785, 1979.

[Fos 84] Foster,M.A. et al. - *Nuclear Magnetic Resonance Pulse Sequence and Discrimination of High- and Low-fat Tissues*. Magnetic Resonance Imaging. 2:187-192, 1984.

[Fuk 72] Fukunaga, K.. - *Introduction to Statistical Pattern Recognition*. Academic Press, London, 1972.

[Goo 85] Gool,L.V.; Dewaele,P; Oosterlink,A. - *Texture Analysis Anno 1983*. Computer Vision, Graphics and Image Processing. 29:336-357, 1985.

[Har 73] Haralick,R.M. at al. - *Textural Features for Image Classification*. IEEE Trans. on Systems, Man and Cybernetics. SMC-3(6):610-621, 1973.

[Loh 88] Loh,H.; Leu,J. - *The Analysis of Natural Textures Using Run Length Features*. IEEE Trans. on Industrial Electronics. 35(2):323-328, 1988.

[Tay 86] Taylor,D.G.; Bushell,M.C. - *NMR Characterisation of Healthy Rat Tissues in vivo*. Clin. Phys. Physiol. Meas. 7(1):57-62, 1986

[The 89] Therrien, C.W. - *Decision, Estimation and Classification: an introduction to pattern recognition and related topics*. John Wiley & Sons, New York, 1989.